Design and Verification of Hybrid Dynamic Worm Detection System

Hongyan Wang, Xiaoguang Tang, Cheng Zhuang, Changqin Fu

1. Hebei Normal University of Science and Technology, Qinghuangdao, China
   wanghy63@163.com
2. Hebei Normal University of Science and Technology, Qinghuangdao, China
   txg60@163.com
3. Hebei Normal University of Science and Technology, Qinghuangdao, China
   chengz2004@126.com
4. Hebei Normal University of Science and Technology, Qinghuangdao, China
   cq_fu@126.com

Abstract

This paper proposes a new design scheme for hybrid dynamic worm detection system. In the proposed scheme, the features of matching detection method based on protocol analysis, Bayesian detection method based on probability, and honeypot-based detection method are integrated for a relative complemented function to develop the advantages and avoid the disadvantages of each detection method. The designed hybrid dynamic worm detection system can be used for effective detection of abnormal data behavior of the hosts attacked by worms, and it also serves for dynamic real-time update of the rule base. Then a preliminary functional verification is performed on this system.

Keywords: Computer Worm, Feature Extraction, Detection of Bayesian Probability, Honeypot

1. Introduction

Worm virus has become a serious threat in internet safety. Warm can spread from one note to another in the network, thus it has great complexity and uncertainty in spreading and action character[1]. Therefore the research of worm detection technology has become one of the hot topics of network safety. The current protective measures include firewall, invasion detection system, virus protection system, etc. These methods can control the attack of worm in some degree, but still there are a lot of limitations in different aspects. They cannot solve network safety problem completely and effectively. According to recent research, using one single detection technology will suffer great limitation. For example, detection technology based on character matching has two fundamental defects[2]. First, it requires extremely large number of computation. Second, it uses a fixed character model to detect the attack, so that only the specific characteristic attack can be detected. If the attack has a slightly variance, it can avoid being detected easily. Therefore, multiple detection methods must be combined together in order to increase the satiability and safety of the network. In addition, new types of worms require the detection system has certain degrees of dynamic adaptive character. This article presents a hybrid dynamic worm detection system based on LAN on the basis of the current network technology.

2. General Architecture of the System

The general design architecture of the system is shown in Figure 1, and the functional modules of the system include:

Worm detection and analysis engine program: It is used to capture the data stream of the network, extract the communication mode between nodes, and, with several detection methods, determine whether the worm attack packet exists in local network;

Packet capture module: It is used to capture and filter packet from the network, and transfer the target data packet to other modules for processing;
Protocol analysis module: It is used to analyze the network data packet transferred from the capture module, and extract the information including the source address, destination address, source port, destination port, protocol type, and data length etc.;

Feature resolution program: It is used to resolve the rules in the feature library and prepares for feature matching detection;

Feature matching module based on protocol analysis: It is used for feature detection via the application of known attack mode;

Bayesian worm detection module based on probability: It is used to detect worm attack by analyzing whether the probability value of first connection failure between the suspicious host and the target host is higher than the worm determination threshold, with simultaneous work of honeypot-based detection module;

Worm detection module based on honeypot technique: Through honeypot analysis, it can access the network packet whose IP cannot be found in local network, which is characterized by bits of IP traffic required for judgment and can be used for worm attack detection at the early stage;

Feature extraction module: It is used to extract the featured character strings of worm attack and add them into the feature library after honeypot-based detection;

Console program: It is used to administrate the running of work detection engine, receive worm alarms from the detection engine, and save the alarm information into the database for future inquiry and operation.

3. Design of Main System Modules

3.1. Packet capture module

The calling procedure of network packet capture function developed on the basis of Libpcap packet capture of API function interface is shown in Figure 2, and all packets to local network can be captured by setting the network card into hybrid mode[3]. Review or processing of the packet can be realized with the callback function (packet_callback).
3.2. Protocol analysis module

Based on TCP/IP analysis principles[4], the basic protocol analysis process can be concluded as shown in Figure 3, and the functionally dependent relation can be identified as shown in Figure 4.

![Diagram of Protocol Analysis Process]

**Figure 3. Protocol Analysis Process**

![Diagram of Functional Relation]

**Figure 4. Functional Relation**

3.3. Feature matching module based on protocol analysis

The feature matching processes can be reached based on protocol analysis techniques: after capture of network packet, call the function of `analysis_ethernet()` to conduct length and LLC check for packet,
and then call the function of analysis_ip() to analyze IP packet head as for those packets complying with Ethernet packet format. As the IP packet monitored may be fragment packets, the information shall be temporarily saved for each for reorganization after all fragments are received. An ipHashTab1 Table, with the structure of ipHashTab, is adopted here for temporary saving of incomplete fragment packets, and their key assignments include the fragment ID field combinations of source address and destination address of the IP packet. At the same time, the upper-layer protocol type is obtained. For TCP, the function of analysis_tcp() is called to get the destination port and source port and save the source and target IP sources, as well as the source and destination ports, of the lower-layer protocol into another tcpHashTab with the structure of HashTab. In this process, it is firstly to search the tcpHashTab and determine the element location to save corresponding key assignments; then to take a node in the linked list, match the featured character strings designated with payloadOptNode linked list, and call the processing function to detect the match. The return result of “true” indicates occurrence of worm attack, and the response layer shall be submitted finally for safety response; while search completed in tcpHashTab or the return result of “false” indicates that no match in all payloadOptNode linked list, i.e. no occurrence of defined worm attack, and the packet shall be given up.

3.4. Worm detection module based on Bayesian Probability

The flow diagram of Bayesian computer worm detection is designed according to the process of Bayesian computer worm detection algorithm, as shown in Figure 5.

![Flow Diagram of Bayesian Computer Worm Detection](image)

Figure 5. Flow Diagram of Bayesian Computer Worm Detection
After sufficient data analysis, the following parameter values are obtained[7]: 
\[ P(\text{failure}|L(x)=w) = 0.8; \quad P(\text{failure}|L(x)=\neg w) = 0.1; \quad \text{worm-attacked host threshold} \quad \alpha = 0.999995; \quad \text{normal host threshold} \quad \beta = 0.0001; \quad \text{and initial probability of worm attack for target host} \quad P(L(x)=w) = 0.5. \]

If the connection is established, the calculation formula is:

\[
P(L(x) = w | \text{success}) = \frac{P(L(x) = w) P(\text{success} | L(x) = w)}{\sum_{j=1}^{2} P(L(x)_j) P(\text{success} | L(x)_j)}
\]

(1)

If the connection is failure, the calculation formula is:

\[
P(L(x) = w | \text{failure}) = \frac{P(L(x) = w) P(\text{failure} | L(x) = w)}{\sum_{j=1}^{2} P(L(x)_j) P(\text{failure} | L(x)_j)}
\]

(2)

3.5. Worm detection and feature extraction module based on honeypot

Worm detection system, with aid of honeypot, captures all suspicious packets sent to hosts in the honeypot network[8], analyzes the suspicious packets and determines whether worms exist therein; in addition, it extracts the featured character strings by automatic extraction algorithm based on worm features, for updating the worm feature library. The basic structure of honeypot network is shown as Figure 6.

Method for worm feature detection and extraction: based on determining whether there is outward connection request in captured packet, firstly obtain the source address and destination port of the packet to work out hash value, and then locate elements in hashTab2. Save the destination address and destination port for connecting outward packet and hashTab2, and search for nodes, in hashTab2 linked list, with the target IP identical with the source IP of captured packets, as well as the same destination port with that of the captured ones. Any packet with such node proves to be outward packet. Then, search the packet linked list of the node and compare the length and content of the packet, to determine the existence of worm attack by judging whether the length of the maximum common substring is beyond the worm threshold. Finally, save the worm features into the feature library, for detection as a known worm via its feature matching.

![Figure 6. Basic Structure of Honeypot](image)

4. Functional Verification of the System
Provided with the laboratory LAN, a general network configuration scheme is established as the application environment of the system, including both the network connection of hardware and the installation and configuration of software.

4.1. Verification of known worm detection function

The verification mainly serves for feature matching worm detection method based on protocol analysis technology. Upon the worm detection module based on feature matching is started, the feature codes of known worms in feature library will be automatically loaded by the system[9]. Choose a host to run the worm attack program, with the host node in local network as the target. The communication data traffic generated by the attack will be captured at the detector. Then feature matching is conducted after protocol analysis of worm detection system, and the detection system will produce alarms according to detection outputs. A part of attack process captured is shown as Figure 7 and 8.

![Figure 7. Attack Execution Process of THCIISLame.exe](image)

![Figure 8. Detection Output of THCIISLame.exe Attack](image)
4.2. Verification of unknown worm detection function

It mainly verifies the realization of Bayesian worm detection method based on probability and honeypot-based worm detection and feature extraction method. Firstly, transform the existing worms with ADMmutate[10] worm generation engine to simulate unknown worms, and the host node in local network and honeypot network can be taken as the attack target. Then perform experimental verification on these two methods respectively. For Bayesian worm detection based on probability, use the worm host with Linux Fedora core 7 operation system in local network to run the simulated worm attack program, and, utilize the function of creating a new process with function of fork() under Linux, to designate different attack targets for each process to test the functional effectiveness of Bayesian worm detection module with the knowledge that a cycling command operation can simultaneously run several simulated worm attack programs in a short time. As regards honeypot-based worm detection method, use the worm host with Linux Fedora core7 operation system to run the simulated worm attack program to attack the host in honeypot networked, and adopt the honeypot-based lightweight worm detection and feature extraction method at the same time to extract the worm features, which are to be displayed at the console. The detection outputs are shown in Figure 9 and 10.

![Figure 9. Detection Output of Ncexpl Attack Based on Honeypot](image1)

![Figure 10. Detection Output of Ncexpl Attack Based on Bayesian Method](image2)
5. Conclusion

This paper presents a hybrid worm detection model based on LAN, designs a worm detection system using proposed model and conducts prototype realization on its key modules. The main designed modules include packet capture module, protocol analysis module, and key modules at detection and analysis layer of this system, i.e. the feature matching module for known worms based on protocol analysis, Bayesian worm detection module for unknown worms based on probability, and worm detection and feature extraction module based on honeypot. Then an application environment is established for functional verification of the system. The final outputs indicate that this system can be used to detect worm attack and extract worm features effectively.

Acknowledgement

The authors are grateful to Science and Technology Bureau of Qinhuangdao for providing the Science and Technology Research and Development Direction Program (ID2006-2-39). The authors would also like to thank the support of Teaching Research Program (ID JYZD20104) from Hebei Normal University of Science and Technology.

6. References